

# Multi-Step Bayesian Optimization for One-Dimensional Feasibility Determination

J. Massey Cashore

Lemuel Kumarga

Peter I. Frazier

## Abstract

Bayesian optimization methods allocate limited sampling budgets to maximize expensive-to-evaluate functions. One-step-lookahead policies are often used, but computing optimal multi-step-lookahead policies remains a challenge. We consider a specialized Bayesian optimization problem: finding the superlevel set of an expensive one-dimensional function, with a Markov process prior. We compute the Bayes-optimal sampling policy efficiently, and characterize the suboptimality of one-step lookahead. Our numerical experiments demonstrate that the one-step lookahead policy is close to optimal in this problem, performing within 98% of optimal in the experimental settings considered.

## 1 Introduction

We consider the problem of adaptively allocating sampling effort to efficiently estimate sub- and super-level sets of a one-dimensional Markov process, or more general additive functionals of this process. We use a decomposition property to show how the optimal procedure may be computed efficiently, circumventing the curse of dimensionality. We then use our ability to compute the optimal policy to study the suboptimality gap of commonly used one-step lookahead procedures in this problem.

The problem we consider falls within the class of problems considered in the large and rapidly growing literature on Bayesian optimization [22, 25, 21, 12, 33], which seeks to develop adaptive sampling algorithms that estimate functionals, especially the location of a global maximum, of some underlying and unknown function in a query efficient way. Such problems arise when optimizing an objective that is computed via a long-running computer code [12, 33] or some other expensive process [4, 15] that severely limits the number of times it may be sampled. This literature places a Bayesian prior distribution on the underlying function, and views it as a realization of a stochastic process, most frequently a Gaussian process.

In such problems, a Bayes-optimal algorithm is one that minimizes the expected loss under the prior suffered from mis-estimation of the underlying functional of interest, where the cost of sampling is either factored directly into the objective (as considered by [10, 3]), or a sampling budget is enforced as a constraint (as considered by [17, 13]). When optimization is the goal, this loss function is the opportunity cost — the difference in value between the point that is believed to be the best, and the value of the true global optimum — but when other functionals are of interest another loss function may be appropriate.

In principle, a Bayes-optimal algorithm may be computed using stochastic dynamic programming by understanding that this problem is a partially observable Markov decision process (POMDP) [17]. However, the curse of dimensionality [29] prevents actually computing the solution through brute-force approaches.

Thus, almost all of the literature has focused on approximate schemes, which in many cases are inspired by this view of the problem as a partially observable Markov decision process, but that do not actually solve the POMDP. Two commonly used methods of this type are the expected improvement method [25, 21] and the knowledge-gradient method [14, 32], which use one-step lookahead approaches, based on different assumptions about what points are eligible for selection once sampling stops [15]. Two-step lookahead approaches have also been implemented computationally in [2, 17].

In contrast, we focus on calculating the Bayes-optimal algorithm. Our primary contribution is to show that it can be computed efficiently in Bayesian optimization problems that satisfy four assumptions:

- the underlying function is one-dimensional, as considered by [22, 6, 7, 9, 8, 27, 23].
- the Bayesian prior on this function has the Markov property (e.g., a Wiener process prior, as used by [22, 27, 1, 38, 30, 24, 23, 9], or an Ornstein-Uhlenbeck prior [26, 28]).

- the loss function is additive across location, as arises when the goal is to determine feasibility of points, as in [16], or to determine the set of points that are better than some known standard, as in [37].
- the limit on sampling is imposed as an additive cost in the objective, as in [10, 3], or as a constraint on the expected number of samples taken, as in [17, 13].

As a second contribution, we also provide an upper bound on the value of the Bayes optimal policy when the limit on sampling is imposed as an almost sure constraint on the number of samples taken.

While one dimensional feasibility determination problems do arise in practice [20], and we expect that the optimal policy can provide a great deal of value in those settings, a large fraction of practical Bayesian optimization problems violate one or more of the assumptions above, because many problems are in more than one dimension, and because non-Markov Gaussian processes are often used as priors [18, 31, 4]. Optimization is also a more common goal in the literature than super-level set determination (though in practice it is often just as useful to provide a set of points that perform well, i.e., that reside in some super-level set, from which a final decision can be selected based on other criteria).

Thus, we view our primary contribution as providing a specialized but nevertheless rich class of Bayesian optimization problems on which the performance of widely applicable heuristic procedures, such as the one-step lookahead procedures described above, may be studied relative to Bayes-optimal procedures. This guides algorithm development — if a heuristic procedure performs close to optimal on a set of problems, then this suggests that further improvement is not necessary even for other similar problems for which the optimality gap cannot be evaluated. In contrast, if all known heuristic procedures perform substantially worse than optimal on a set of problems, then this suggests that further algorithm development is worthwhile.

There is some complementary theoretical analysis in the literature of Bayes-optimal procedures for this and related problems. Much of it focuses on asymptotic analyses, and includes proofs of consistency for the Efficient Global Optimization (EGO) [35] and P algorithms [6], as well as convergence rates for these and closely related algorithms [7, 5, 9]. In terms of finite-time analyses, [34, 19] provide regret bounds for the closely related problem of Bayesian optimization in the bandit setting, but while these bounds characterize performance, slack in the bounds' constants creates a potentially large multiplicative gap in which performance may lie. In the problems of multiple comparisons with a known standard and stochastic root-finding, procedures for computing explicit Bayes optimal procedures have been developed [37, 36], but these problems are only distantly related to Bayesian optimization. Thus, exact performance of optimal finite-time procedures has remained unknown in Bayesian optimization.

Below, in Section 2, we provide a formal description of the problem. Our main results are in Section 3, where we significantly reduce the state-space for a dynamic program giving rise to a Bayes-optimal policy. In Section 4 we consider the relationship between the cost-per-sample setting and the constrained-budget setting, showing how the optimal value for the former can be used to compute the optimal value of the latter. In Section 5 we present numerical results, illustrating the behavior of the optimal policy and using it to analyze the optimality gap for a one-step lookahead procedure. Finally, in Section 6, we conclude.

## 2 Problem Description

Let  $Y = (Y(x) : x \geq 0)$  be a Markov process over the positive real line, and let  $[a, b]$  be a given interval,  $0 < a < b < \infty$ . We consider adaptive sampling policies that characterize  $Y$  over  $[a, b]$ .

We will consider histories of the form  $\{(x_t, Y(x_t)) : t = 1, \dots, T\}$  for some sequence of adaptively chosen points  $(x_t : t = 1, \dots, T)$  at which measurements occur.

Let  $\mathcal{H} = \cup_{T=0}^{\infty} (\mathbb{R}_+ \times \mathbb{R})^T$  be the space of all possible histories. A policy  $\pi : \mathcal{H} \mapsto \mathbb{R}_+ \cup \{\Delta\}$  is a measurable function that maps the current history to either a point to be sampled next, or to the symbol  $\Delta$ , which indicates the decision to stop. We let  $\Pi$  indicate the space of all such policies.

We begin with an initial history  $H_0 \in \mathcal{H}$ . For simplicity of analysis we assume that  $H_0$  contains endpoint observations, that is  $(a, Y(a)), (b, Y(b)) \in H_0$ , but our results can be extended to the case where it does not. We define histories  $H_t$  and decisions  $x_t$  recursively, letting  $x_{t+1} = \pi(H_t)$  and letting

$$H_{t+1} := \begin{cases} H_t \cup \{(x_{t+1}, Y(x_{t+1}))\}, & \text{if } x_{t+1} \neq \Delta, \\ H_t, & \text{if } x_{t+1} = \Delta, \end{cases} \quad (2.1)$$

so that the point sampled and the resulting observation of  $Y$  is added to the history if the policy chooses to sample, and the history remains unchanged once the policy chooses to stop sampling. As indicated, if a policy measures a point already in the history, then the history remains unchanged. Also note that because the history is a set of tuples, the policy's next action cannot depend on the order in which observations were made.

We define  $\tau = \inf \{t \geq 0 : \pi(H_t) = \Delta\}$  to be the total number of samples taken by a policy. When necessary, we will write  $\tau^\pi$  to emphasize the policy on which  $\tau$  depends.

We also define  $\mathbb{P}^\pi$  to be the distribution over histories with respect to the randomness in  $Y$  and the decisions made by  $\pi$ , for any  $\pi \in \Pi$ . We let  $\mathbb{E}^\pi$  denote the expectation with respect to this distribution.

We seek to characterize  $Y$  by assigning each point  $x \in [a, b]$  a label, or class, based on our knowledge of  $Y(x)$ . Suppose there are  $n < \infty$  classes to which a point might belong and let  $I$  be an index set such that each element corresponds to one class. At time  $\tau$ , we will use the information collected, encoded in  $H_\tau$ , to classify each point in  $[a, b]$ : based on  $H_\tau$  we construct a partition  $\{B_i : i \in I\}$  of  $[a, b]$ , such that each  $B_i$  is a measurable subset of  $[a, b]$ . If  $x \in B_i$ , we say that  $x$  belongs to the  $i$ th class. We will receive a reward  $R_{[a,b]}(H_\tau)$ , defined below, that depends on the accuracy of this classification.

To formalize this we first choose bounded measurable functions  $f_i : \mathbb{R} \rightarrow \mathbb{R}$  for each  $i \in I$ . The function  $f_i$  is meant to reward or penalize the classification  $x \in B_i$  given the true value  $Y(x)$ . In addition to requiring the  $f_i$  be measurable and bounded, we also require that, for each  $i \in I$ ,  $f_i$  satisfies the following inequality:

$$\int_{[a,b]} \mathbb{E}[|f_i \circ Y(x)|] dx < \infty. \quad (2.2)$$

By Fubini's theorem, this inequality will allow us below to interchange the integral over the domain of  $Y$  and the expectation over the randomness in  $Y$ .

Now, fix any partition  $\mathbb{B} = \{B_i : i \in I\}$  and  $H \in \mathcal{H}$ . We define the expected reward of the partition  $\mathbb{B}$  given  $H_\tau = H$  over  $[a, b]$  to be

$$R_{[a,b]}(H, \mathbb{B}) = \mathbb{E} \left[ \sum_{i \in I} \int_{B_i} f_i \circ Y(x) dx \mid H \right] = \sum_{i \in I} \int_{B_i} \mathbb{E}[f_i \circ Y(x) \mid H] dx, \quad (2.3)$$

where the last equality holds due to (2.2) and Fubini's theorem. Observing (2.3), a partition maximizing  $R_{[a,b]}(H, \mathbb{B})$  is any  $\mathbb{B}^* = \{B_i^* : i \in I\}$  such that for all  $j \in I$ , if  $x \in B_j$ , then

$$j \in \operatorname{argmax}_i \mathbb{E}[f_i \circ Y(x) \mid H].$$

That is,  $x$  belongs to any class  $j$  maximizing  $\mathbb{E}[f_j \circ Y(x) \mid H]$ .

We define the expected reward  $R_{[a,b]}(H)$  for any  $H \in \mathcal{H}$  to be the expected reward of any optimal partition given  $H$ . That is,

$$R_{[a,b]}(H) = R_{[a,b]}(H, \mathbb{B}^*) = \int_{[a,b]} \max_i \mathbb{E}[f_i \circ Y(x) \mid H] dx. \quad (2.4)$$

Note that because we choose the functions  $\{f_i : i \in I\}$  to be bounded, it follows that there exists some constant  $C$  such that

$$|R_{[a,b]}(H)| \leq C(b - a), \quad (2.5)$$

for every  $H \in \mathcal{H}$ .

We now describe how this framework can be specialized to the problem of estimating superlevel sets. Recall the superlevel set of a function  $g : [a, b] \rightarrow \mathbb{R}$  with respect to the threshold  $k$  is the set  $\{x \in [a, b] : g(x) \geq k\}$ . In this context we use the index set  $I = \{+, -\}$  corresponding to the classification of a point as above or below the threshold. We give two reasonable choices for the functions  $f_+$  and  $f_-$ :

1.  $f_+(y) = \mathbb{1}\{y \geq k\}$  and  $f_-(y) = \mathbb{1}\{y \leq k\}$ . These functions are clearly measurable, bounded, and satisfy the inequality (2.2).

2.

$$f_+(y) = \begin{cases} y - k & \text{if } |y - k| \leq C, \\ C \cdot \text{sign}(y - k) & \text{otherwise,} \end{cases} \quad \text{and} \quad f_-(y) = \begin{cases} k - y & \text{if } |k - y| \leq C, \\ C \cdot \text{sign}(k - y) & \text{otherwise,} \end{cases}$$

for some constant  $C > 0$  chosen a priori. We only consider these reward functions for Markov processes  $Y$  such that the inequality (2.2) is satisfied.

Although we focus on superlevel set detection, this framework can be used to classify points based on other properties of  $Y(x)$ . For example, we could consider two thresholds for the range of  $Y(x)$ , and classify each point as being below both, above both, or in between them.

The performance of a policy  $\pi$  at state  $H$  over  $[a, b]$  is the expected value of the final reward less the cost associated with the expected number of samples starting from an initial history  $H$ . That is, given a cost-per-sample of  $c > 0$ , the performance is defined as:

$$\text{Per}(\pi, c, H) = \mathbb{E}^\pi [R_{[a,b]}(H_\tau) - c\tau \mid H]. \quad (2.6)$$

As a consequence of (2.5), a policy that has non-zero probability of taking infinitely many samples at a state  $H$  (i.e.  $\mathbb{P}^\pi(\tau = \infty \mid H) \neq 0$ ) achieves a performance of  $-\infty$  at  $H$ .

Finally, the value of a state  $H$  is defined to be the supremum of the performance over all policies. That is,

$$V_{[a,b]}(H) = \sup_{\pi \in \Pi} \mathbb{E}^\pi [R_{[a,b]}(H_\tau) - c\tau \mid H]. \quad (2.7)$$

We call (2.7) the cost-per-sample setting. Below, in section 4, we consider two other related settings: budget-constrained and expected-budget-constrained.

Now, in Section 3, we show how to approximately compute  $\epsilon$ -optimal policies for the cost-per-sample setting when summability is satisfied, i.e., one that attains within  $\epsilon$  the supremum in (2.7) for the initial history  $H_0$  and any  $\epsilon > 0$ . We refer to such an optimal policy as " $\epsilon$ -Bayes-optimal" because it is  $\epsilon$ -optimal with respect to an expectation taken over the probability distribution of  $Y$ , which can be understood to be a Bayesian prior distribution.

### 3 Main Results for Cost-Per-Sample Case

To compute a Bayes-optimal policy, we focus on efficiently computing the value function defined in (2.7). Naive dynamic programming can be used, but the dimensionality of the portion of the state space reachable after  $t$  samples grows linearly in  $t$ , causing the volume of the state space, and thus the memory and computation required for dynamic programming, to grow exponentially. Our main result decomposes the value function, showing that it is completely determined by its value on some 4-dimensional set, leading to its computation as a tractable dynamic program over a state space of small constant dimension. In particular, we prove the following:

**Theorem 1.** *Fix interval  $[a, b]$  and let  $H \in \mathcal{H}$  be such that observations of  $a$  and  $b$  are included. Let  $x_1, \dots, x_{t+2}$  be the observations in  $H$  contained within  $[a, b]$ . Suppose they are ordered such that  $x_i < x_{i+1}$  for all  $1 \leq i < t + 2$ , so  $x_1 = a$  and  $x_{t+2} = b$ . Define  $H_i = \{(x_i, y_i), (x_{i+1}, y_{i+1})\}$  for each  $1 \leq i < t + 2$ . Then*

$$V_{[a,b]}(H) = \sum_{i=1}^{t+1} V_{[x_i, x_{i+1}]}(H_i). \quad (3.1)$$

The importance of this theorem is that  $V_{[a,b]}(H)$  is completely determined by its values on  $\{H \in \mathcal{H} : |H| = 2\}$ , greatly reducing the effective dimension of the dynamic program's state space. We show below how this dimension reduction can be used in a recursive algorithm over a 4-dimensional state space (the set of histories of length two) to find  $V_{[a,b]}(H)$ . Recall that knowledge of the value function  $V_{[a,b]}$  at every state can lead to  $\epsilon$ -optimal policies. Indeed, if  $\epsilon = \kappa_1 + \kappa_2 + \dots$ , and  $\pi$  is a policy that, at the  $t$ th step, selects a point  $x_t$  to sample that is within  $\kappa_t$  of the optimal, then  $\pi$  is  $\epsilon$ -optimal. The point  $x_t$  can be any point such that  $V_{[a,b]}(H_t) - \kappa_t \leq \mathbb{E}[V_{[a,b]}(H_t \cup \{(x_t, Y(x_t))\}) \mid H_t]$ . For further details, see [11] section 5.

We can further reduce the state space when  $Y$  satisfies additional structure. For any  $\ell \in \mathbb{R}$ , define the shift operator  $T_\ell : \mathbb{R} \rightarrow \mathbb{R}$  by  $T_\ell(x) = x + \ell$ . We will apply  $T_\ell$  to elements of  $\mathcal{H}$ , and adopt the convention that  $T_\ell(H) = \{(x + \ell, y) : (x, y) \in H\}$ , i.e.  $T_\ell$  only translates the location of the observations in  $H$ , and not their values. We say the Markov process  $Y$  is *translation invariant* if, for any  $H \in \mathcal{H}$ ,  $y \in \mathbb{R}$ ,  $x \in \mathbb{R}_+$  and  $\ell \in \mathbb{R}$  such that  $x + \ell \geq 0$ ,

$$\mathbb{P}(Y(x) \in dy \mid H) = \mathbb{P}(Y(x + \ell) \in dy \mid T_\ell(H)). \quad (3.2)$$

The following proposition establishes that if the Markov process is translation invariant, so is the value function.

**Proposition 2.** *Suppose  $Y$  is translation invariant. Fix interval  $[a, b]$  and pick any  $\ell \in \mathbb{R}$  such that  $a + \ell \geq 0$ . Pick any history  $H \in \mathcal{H}$  and let  $H' = T_\ell(H)$ . Then*

$$V_{[a, b]}(H) = V_{[a', b']}(H') \quad (3.3)$$

where  $a' = a + \ell$  and  $b' = b + \ell$ .

Thus when  $Y$  satisfies translation invariance, the value function is completely determined by its values on  $\{H \in \mathcal{H} : |H| = 2, (0, y_0) \in H\}$ . (The choice of 0 in the  $(0, y_0) \in H$  condition is arbitrary; one may replace 0 by any other constant in the domain of  $Y$ ). In this case,  $V_{[a, b]}$  can be computed as the result of a dynamic-programming-like recursion over a 3-dimensional, rather than 4-dimensional, state space, as described below. This reduction in dimension enables faster computation with less memory.

To prove our main results, we first state two technical lemmas:

**Lemma 3.** *Let  $H \in \mathcal{H}$  contain  $t$  observations. Let  $H^I = \{(x, y) \in H : x \in [a, b]\}$  denote the set of initial observations inside  $[a, b]$ . Then*

$$V_{[a, b]}(H) = V_{[a, b]}(H^I). \quad (3.4)$$

For the following lemma and rest of this section we adopt the notation that, for  $H \in \mathcal{H}$  and  $A \subseteq \mathbb{R}$ ,  $H \cap A = \{(x, y) \in H : x \in A\}$ . We will also write  $x \in H$  to mean there exists  $y \in \mathbb{R}$  such that  $(x, y) \in H$ .

**Lemma 4.** *Fix interval  $[a, b]$ . Let*

- $\Pi^1 = \{\pi \in \Pi : \mathbb{P}^\pi(\tau < \infty \mid H) = 1 \ \forall H \in \mathcal{H}\}$  be the set of policies that almost surely take finitely many samples.
- $\Pi^2_{[a, b]} = \{\pi \in \Pi : \pi(H) \in [a, b] \ \forall H \in \mathcal{H}\}$  be the set of policies that only take samples in  $[a, b]$ .
- $\Pi^3 = \{\pi \in \Pi : \pi(H) \neq x \text{ if } x \in H \ \forall H \in \mathcal{H}\}$  be the set of policies that do not sample the same point twice.

Let  $\bar{\Pi}_{[a, b]} = \Pi^1 \cap \Pi^2_{[a, b]} \cap \Pi^3$ . For all  $H \in \mathcal{H}$ , define

$$\bar{V}_{[a, b]}(H) = \sup_{\pi \in \bar{\Pi}_{[a, b]}} \mathbb{E}^\pi [R_{[a, b]}(H_\tau) - c\tau \mid H]. \quad (3.5)$$

Then  $V_{[a, b]}(H) = \bar{V}_{[a, b]}(H)$  for all  $H \in \mathcal{H}$ .

Lemma 3 says that if  $H \in \mathcal{H}$  contains endpoint observations then the only points in  $H$  that affect  $V_{[a, b]}(H)$  are those within  $[a, b]$ . Lemma 4 constructs a subset of  $\Pi$  containing an optimal policy. The proofs of the above lemmas, as well as Proposition 2 are contained in the appendix. We are now in a position to prove the main decomposition theorem.

**Proof of Theorem 1.** We proceed by induction on  $t$ . When  $t = 0$  the summation contains only one term and the result is established. Fix  $t > 0$  and suppose the decomposition (3.1) holds for any  $|H| < t + 2$ .

Define  $\tau_A$  with respect to any policy  $\pi$  to be the number of points  $\pi$  chooses to sample inside the set  $A$ , for some  $A \subseteq [a, b]$ . Thus if  $\{w_i : 1 \leq i \leq \tau\}$  is the set of points sampled by  $\pi$ ,  $\tau_A = \sum_{i=1}^\tau \mathbb{1}\{w_i \in A\}$ . By Lemma 4 we restrict our attention to  $\pi \in \bar{\Pi}_{[a, b]}$ . In particular if  $(x, Y(x)) \in K$ ,  $\pi$  will not choose

to sample at  $x$  again given initial state  $K$ . Thus  $\tau_{[a,b]} = \tau_{[a,x]} + \tau_{[x,b]}$  conditioned on any initial history containing  $(x, Y(x))$ . This is because the only point in  $[a, x] \cap [x, b]$  is  $\{x\}$ , and the lone sample of  $\{x\}$  is in the initial history and it is not counted in  $\tau_{[a,x]}$  or  $\tau_{[x,b]}$ . From the Markov property, it is also clear that  $R_{[a,b]}(H_\tau) = R_{[a,x]}(H_\tau) + R_{[x,b]}(H_\tau)$  almost surely conditioned on any initial history containing  $x$ .

Now, fix some  $1 < i < t + 2$ , so that  $x_i$  is not  $a$  or  $b$ . Note that

$$V_{[a,b]}(H) = \sup_{\pi \in \bar{\Pi}_{[a,b]}} \mathbb{E}^\pi [R_{[a,b]}(H_\tau) - c\tau_{[a,b]} \mid H] \quad (3.6)$$

$$= \sup_{\pi \in \bar{\Pi}_{[a,b]}} (\mathbb{E}^\pi [R_{[a,x_i]}(H_\tau) - c\tau_{[a,x_i]} \mid H] + \mathbb{E}^\pi [R_{[x_i,b]}(H_\tau) - c\tau_{[x_i,b]} \mid H]) \quad (3.7)$$

$$\leq \sup_{\pi \in \bar{\Pi}_{[a,b]}} \mathbb{E}^\pi [R_{[a,x_i]}(H_\tau) - c\tau_{[a,x_i]} \mid H] + \sup_{\sigma \in \bar{\Pi}_{[a,b]}} \mathbb{E}^\sigma [R_{[x_i,b]}(H_\tau) - c\tau_{[x_i,b]} \mid H] \quad (3.8)$$

$$= \sup_{\pi \in \bar{\Pi}_{[a,x_i]}} \mathbb{E}^\pi [R_{[a,x_i]}(H_\tau) - c\tau_{[a,x_i]} \mid H] + \sup_{\sigma \in \bar{\Pi}_{[x_i,b]}} \mathbb{E}^\sigma [R_{[x_i,b]}(H_\tau) - c\tau_{[x_i,b]} \mid H] \quad (3.9)$$

$$= V_{[a,x_i]}(H) + V_{[x_i,b]}(H). \quad (3.10)$$

The equality between (3.6) and (3.7) holds because  $x_i \in H$ . The equality between (3.8) and (3.9) holds because  $\bar{\Pi}_{[a,x_i]} \subseteq \bar{\Pi}_{[a,b]}$  and Lemma 4 shows the supremum is achieved in  $\bar{\Pi}_{[a,x_i]}$  and similarly for  $\bar{\Pi}_{[x_i,b]}$ . The equality between (3.9) and (3.10) holds because  $\tau_{[a,x_i]}^\pi = \tau^\pi$  for any  $\pi \in \bar{\Pi}_{[a,x_i]}$ .

We now show that  $V_{[a,b]}(H) \geq V_{[a,x_i]}(H) + V_{[x_i,b]}(H)$ . Let  $\pi \in \bar{\Pi}_{[a,x_i]}$  and  $\sigma \in \bar{\Pi}_{[x_i,b]}$ . Define the policy  $\gamma$  by

$$\gamma(H) = \begin{cases} \pi(H \cap [a, x_i]), & \text{if } \pi(H \cap [a, x_i]) \neq \Delta, \\ \sigma(H \cap [x_i, b]), & \text{otherwise.} \end{cases} \quad (3.11)$$

That is,  $\gamma$  is the policy that executes  $\pi$  with input from  $[a, x_i]$  until  $\pi$  chooses to stop sampling, and then executes  $\sigma$  with input from  $[x_i, b]$  until  $\sigma$  chooses to stop sampling. Since  $\pi \in \bar{\Pi}_{[a,x_i]}$  and  $\sigma \in \bar{\Pi}_{[x_i,b]}$  we know  $\tau^\pi$  and  $\tau^\sigma$  are almost surely finite, so  $\gamma$  will fully execute both  $\pi$  and  $\sigma$ . Observe the expected performance under  $\gamma$  is

$$\mathbb{E} [R_{[a,b]}(H_\tau^\gamma) - c\tau_{[a,b]}^\gamma \mid H] = \mathbb{E} [R_{[a,x_i]}(H_\tau^\gamma) - c\tau_{[a,x_i]}^\gamma \mid H] + \mathbb{E} [R_{[x_i,b]}(H_\tau^\gamma) - c\tau_{[x_i,b]}^\gamma \mid H] \quad (3.12)$$

$$= \mathbb{E} [R_{[a,x_i]}(H_\tau^\pi) - c\tau_{[a,x_i]}^\pi \mid H] + \mathbb{E} [R_{[x_i,b]}(H_\tau^\sigma) - c\tau_{[x_i,b]}^\sigma \mid H]. \quad (3.13)$$

where the decomposition  $\tau_{[a,b]}^\gamma = \tau_{[a,x_i]}^\gamma + \tau_{[x_i,b]}^\gamma$  holds under  $\gamma$  because  $(x_i, Y(x_i)) \in H$  and so  $\gamma$  never samples  $x_i$ . Thus  $V_{[a,b]}(H) \geq V_{[a,x_i]}(H) + V_{[x_i,b]}(H)$ . As we have already established the reverse inequality, it follows that  $V_{[a,b]}(H) = V_{[a,x_i]}(H) + V_{[x_i,b]}(H)$ .

Now, we partition  $H$  about  $x_i$ : Let  $H_{\leq i} = \{(x, w) \in H : x \leq x_i\}$  and  $H_{\geq i} = \{(x, w) \in H : x \geq x_i\}$ . By Lemma 3,  $V_{[a,x_i]}(H) = V_{[a,x_i]}(H_{\leq i})$  and  $V_{[x_i,b]}(H) = V_{[x_i,b]}(H_{\geq i})$ . Hence  $V_{[a,b]}(H) = V_{[a,x_i]}(H_{\leq i}) + V_{[x_i,b]}(H_{\geq i})$ . However, since  $x_i$  was chosen to not be an endpoint,  $|H_{\leq i}| < t + 2$  and  $|H_{\geq i}| < t + 2$ . Thus by the induction hypothesis,

$$V_{[a,b]}(H) = \sum_{j=1}^{i-1} V_{[x_j, x_{j+1}]}(H_j) + \sum_{j=i}^t V_{[x_j, x_{j+1}]}(H_j) \quad (3.14)$$

$$= \sum_{j=1}^t V_{[x_j, x_{j+1}]}(H_j), \quad (3.15)$$

and the induction holds.  $\square$

Theorem 1 and Proposition 2 give rise to an efficient algorithm for computing the value function, summarized in Algorithm 1. Algorithm 1 takes a discretization  $\{x_1, \dots, x_m\}$  of the domain of  $Y$  and discretization  $\{y_1, \dots, y_n\}$  of the range of  $Y$  as parameters. It returns the 3-dimensional array  $V[y_L, y_R, x]$  over  $y_L, y_R \in \{y_1, \dots, y_n\}$  and  $x \in \{x_1, \dots, x_m\}$ . Each element  $V[y_L, y_R, x]$  of  $V$  is an approximation to  $V_{[0,x]}(H)$  where  $H = \{(0, y_L), (x, y_R)\}$ . Recall Theorem 1 establishes that  $V_{[a,b]}(H)$  is completely determined by its values on  $\{H \in \mathcal{H} : |H| = 2\}$  and by assuming translation invariance Proposition 2 establishes that if  $|H| = 2$

---

**Algorithm 1** Algorithm for computing the value function. Note the computation on line 8 is possible because  $V[w_1, w_2, x'']$  will already be stored for all  $w_1, w_2 \in \{y_1, \dots, y_n\}$  and  $x'' \in \{x_1, \dots, x'_m\}$ . We assume that  $Y$  satisfies translation invariance, and thus Proposition 2 applies.

---

**Require:** Interval length  $\ell$ ,  $Y$ -range discretization  $y_1, \dots, y_n$ ,  $[0, \ell]$ -domain discretization  $x_1, \dots, x_m$ .

**Ensure:**  $x_1 = 0$  and  $x_m = \ell$ .

```

1: for  $y_L = y_1, \dots, y_n$  do
2:   for  $y_R = y_1, \dots, y_n$  do
3:     for  $x = x_1, \dots, x_m$  do
4:       if  $x = 0$  then
5:          $V[y_L, y_R, x] \leftarrow 0$ 
6:       else
7:         Let  $H = \{(0, y_L), (x, y_R)\}$ 
8:          $V[y_L, y_R, x] \leftarrow \max \{ \mathbb{E}[V_{[0,x]}(H \cup \{(x', Y(x'))\} \mid H] - c : x' = x_1, \dots, x\} \cup \{R_{[0,x]}(H)\}$ 
9:       end if
10:    end for
11:  end for
12: end for
13: return  $V$ 

```

---

then we can assume the leftmost observation in  $H$  is at 0. Thus the information in the array  $V$  can be used to approximate  $V_{[a,b]}(H)$  for any  $H$ . (Note that finer discretizations lead to more accurate approximations). If  $Y$  does not satisfy translation invariance only a small modification to Algorithm 1 is needed:  $V$  would have to be a 4-dimensional array, adding one more dimension for the leftmost observation.

The crux of the computation appears on line 8. Theorem 1 establishes that

$$V_{[0,x]}(\{(0, y_L), (x, y_R), (x', Y(x'))\}) = V_{[0,x']}(\{(0, y_L), (x', Y(x'))\}) + V_{[x',x]}(\{(x', Y(x')), (x, y_R)\}).$$

The expectation of the quantities on right side of the above formula will have already been stored by the algorithm for each point in the  $Y$ -range discretization, and so the expectation of the left hand side can be estimated by summing over the  $Y$ -range discretization.

## 4 Upper Bound on the Budget-Constrained Problem

So far in this paper we have considered the cost-per-sample scenario, where the policy may choose how many samples to make without any additional constraints. In this section, we show how the cost-per-sample problem relates to the budget-constrained problem, in which the number of samples the policy can take is constrained.

We first introduce the notion of a randomized policy. Let  $\Pi_R = \{\pi : [0, 1] \times \mathcal{H} \rightarrow \mathbb{R}_+ \cup \{\Delta\}\}$ , that is, the set of policies which take an additional argument inside  $[0, 1]$ . For such policies, we adopt the convention that histories are still updated according to (2.1), with the modification that  $x_{t+1} = \pi(U, H_t)$  where  $U \sim \text{Uniform}([0, 1])$  is drawn once at time 0 and held fixed over time. We call these randomized policies because they may take different actions depending on the random variable  $U$ . We will often write  $\pi(H)$  instead of  $\pi(U, H)$  when it is clear that  $\pi$  is randomized. Note that taking the supremum in equation (2.7) over the larger set  $\Pi_R$  instead of  $\Pi$  does not affect the optimal value, because the deterministic  $\epsilon$ -optimal policies we construct based on Theorem 1 remain  $\epsilon$ -optimal.

Throughout this section we hold an interval  $[a, b]$  fixed. At any state  $H \in \mathcal{H}$  and  $T > 0$ , we define the following sets of constrained policies:

$$\Pi_1(H, T) = \{\pi \in \Pi_R : \mathbb{E}^\pi[\tau \mid H] = T\} \quad \text{and} \quad \Pi_2(H, T) = \{\pi \in \Pi_R : \mathbb{P}^\pi(\tau = T \mid H) = 1\}. \quad (4.1)$$

The policies in  $\Pi_1$  are referred to as expected-budget-constrained policies and the policies in  $\Pi_2$  are referred to as the set of budget-constrained policies. The corresponding value functions are defined as:

$$V_1(H, T) = \sup_{\pi \in \Pi_1(H, T)} \mathbb{E}^\pi[R_{[a,b]}(H_\tau) \mid H] \quad \text{and} \quad V_2(H, T) = \sup_{\pi \in \Pi_2(H, T)} \mathbb{E}^\pi[R_{[a,b]}(H_\tau) \mid H]. \quad (4.2)$$

Budget-constrained-policies are common in practice - it is sometimes easier to allocate a predetermined number of samples than to determine a suitable cost, as the cost-per-sample case requires. Note the above are defined without a cost. This is because  $\mathbb{E}^\pi[\tau | H] = T$  for any  $\pi \in \Pi_1(H, T) \cup \Pi_2(H, T)$ , so any cost term would be constant and not affect the optimal solution.

For the rest of this section we will write the cost-per-sample value function, as defined in equation (2.7), as a function of both the state and the cost. That is, let  $V(H, \lambda)$  indicate  $V_{[a,b]}(H)$  with a cost of  $\lambda$ . Now observe that for any  $H \in \mathcal{H}$ ,  $\lambda > 0$ ,  $T > 0$ ,

$$V_1(H, T) = \sup_{\pi \in \Pi_1(H, T)} \mathbb{E}^\pi [R_{[a,b]}(H_\tau) - \lambda(\tau - T) | H] \quad (4.3)$$

$$\leq \sup_{\pi \in \Pi_R} \mathbb{E}^\pi [R_{[a,b]}(H_\tau) - \lambda\tau | H] + \lambda T \quad (4.4)$$

$$= \sup_{\pi \in \Pi} \mathbb{E}^\pi [R_{[a,b]}(H_\tau) - \lambda\tau | H] + \lambda T \quad (4.5)$$

$$= V(H, \lambda) + \lambda T. \quad (4.6)$$

where the inequality between (4.3) and (4.4) holds because  $\Pi_1(H, T) \subseteq \Pi_R$ , and the inequality between (4.4) and (4.5) holds because the supremum is attained by a non-randomized policy. Thus it follows that for any  $H \in \mathcal{H}$ ,

$$V_2(H, T) \leq V_1(H, T) \leq \inf_{\lambda} V(H, \lambda) + \lambda T \quad (4.7)$$

where the first inequality holds because  $\Pi_2(H, T) \subseteq \Pi_1(H, T)$ . Theorem 5 will establish that the second inequality above is tight, under appropriate assumptions on  $T$ .

We now introduce some notation. For any  $H \in \mathcal{H}$  and  $\pi \in \Pi_R$  let

$$r(\pi, H) = \mathbb{E}^\pi [R_{[a,b]}(H_\tau) | H] \quad \text{and} \quad t(\pi, H) = \mathbb{E}^\pi [\tau | H]. \quad (4.8)$$

For any  $\lambda > 0$  and  $\epsilon > 0$ , let

$$\Pi_\epsilon^*(\lambda) = \{\pi \in \Pi_R : \forall H \in \mathcal{H}, \text{Per}(\pi, \lambda, H) \geq V_{[a,b]}(H, \lambda) - \epsilon\}$$

denote the set of randomized policies whose performance at every state in  $\mathcal{H}$  with a cost of  $\lambda$  is at least  $\epsilon$ -optimal. Let

$$s(\lambda, H) = \limsup_{\epsilon \rightarrow 0^+} \{t(\pi, H) : \pi \in \Pi_\epsilon^*(\lambda)\} \quad (4.9)$$

denote the limit as  $\epsilon$  decreases to 0 of the maximum expected number of samples an  $\epsilon$ -optimal policy takes at any state  $H$  given a cost of  $\lambda$ . The function  $s(\lambda, H)$  will be instrumental in characterizing the policies in  $\Pi_1(H, T)$  and thus the expected-budget-constrained value function  $V_1(H, T)$ . Finally, we define

$$\bar{T}(H) = \lim_{\lambda \rightarrow 0^+} s(\lambda, H). \quad (4.10)$$

Intuitively,  $\bar{T}(H)$  denotes the maximum number of samples a sensible policy takes as the cost of taking samples decreases to 0. Note that  $\bar{T}(H)$  inherently depends on the Markov process  $Y$ . It is natural to expect that as the cost decreases to 0 optimal policies will begin to take more and more samples, and hence  $\bar{T}(H) = \infty$ . However, it is possible to construct Markov processes  $Y$  that are completely characterized by finitely many samples within  $[a, b]$ .

The main result of this section is stated in the following theorem.

**Theorem 5.** *Fix a state  $H_0 \in \mathcal{H}$  and let  $0 \leq T < \bar{T}(H_0)$ . Then*

$$V_1(H_0, T) = \inf_{\lambda} V(H_0, \lambda) + \lambda T.$$

Theorem 5 is nice because it gives an alternate characterization of  $V_1(H, T)$ , but its main importance comes in noting the following:

$$\begin{aligned} V(H, \lambda) + \lambda T &= \sup_{\pi \in \Pi} r(\pi, H) - \lambda t(\pi, H) + \lambda T \\ &= \sup_{\pi \in \Pi} r(\pi, H) + \lambda(T - t(\pi, H)). \end{aligned}$$



The fact that  $V(H, \lambda)$  is a supremum over linear functions of  $\lambda$  implies that it is a convex function of  $\lambda$ . Furthermore, for each value of  $\lambda$  the quantity  $V(H, \lambda) + \lambda T$  can be computed using simulations estimate the performance of the policy defined in Algorithm 1. Thus the actual value of  $V_1(H)$  can be computed as the solution to a convex program in  $\lambda$ , which can be solved easily by algorithms such as bisection search.

**Proof of Theorem 5.** We hold  $H_0 \in \mathcal{H}$  fixed throughout the proof, and for simplicity omit  $H$  from the notation defined above (e.g. we refer to  $s(\lambda, H_0)$  as  $s(\lambda)$ ).

We first note two properties concerning  $s(\lambda)$ :

1.  $\lim_{\lambda \rightarrow \infty} s(\lambda) = 0$ . From inequality (2.5) we know there exists some constant  $C > 0$  such that  $|R_{[a,b]}(H_0)| \leq C$ . Thus the policy that takes 0 samples achieves a performance of  $-C$  at worst. Take  $\lambda > 2C$ . Then the best performance a policy that takes 1 or more samples can achieve is worse than  $C - 2C = -C$ , meaning that the 0-sample-policy is best for such  $\lambda$ .
2.  $s(\lambda)$  is monotonically decreasing. Fix  $\epsilon > 0$ , let  $\lambda < \lambda'$  and pick any  $\pi \in \Pi_\epsilon^*(\lambda)$  and  $\pi' \in \Pi^*(\lambda')$ . By the  $\epsilon$ -optimality of  $\pi$  with respect to  $\lambda$  and  $\pi'$  with respect to  $\lambda'$ , the following inequalities hold:

$$r(\pi) - \lambda t(\pi) \geq r(\pi') - \lambda t(\pi') - \epsilon \quad (4.11)$$

$$r(\pi') - \lambda' t(\pi') \geq r(\pi) - \lambda' t(\pi) - \epsilon. \quad (4.12)$$

Subtracting (4.12) from (4.11) it follows that  $(\lambda' - \lambda)t(\pi) + \epsilon \geq (\lambda' - \lambda)t(\pi') - \epsilon$ . Since  $\lambda' > \lambda$ , it follows that

$$t(\pi) \geq t(\pi') - \frac{2\epsilon}{\lambda' - \lambda}. \quad (4.13)$$

Taking  $\epsilon \rightarrow 0^+$ , we conclude  $s(\lambda) \geq s(\lambda')$ .

Since  $s(\lambda)$  converges to  $\bar{T}(H_0)$  as  $\lambda \rightarrow 0^+$ , converges to  $\infty$  as  $\lambda \rightarrow \infty$ , is monotonically decreasing, and  $0 \leq T < \bar{T}(H_0)$ , it follows that there exists some  $\lambda^*$  such that either  $s(\lambda^*) = T$  or there is a jump discontinuity at  $\lambda^*$  around  $T$ , that is  $\lim_{\lambda \rightarrow \lambda^*+} s(\lambda) \leq T$  and  $\lim_{\lambda \rightarrow \lambda^*-} s(\lambda) \geq T$ . Pick any sequence  $(\bar{\lambda}_n, \underline{\lambda}_n)$  such that  $(\bar{\lambda}_n)$  is decreasing in  $n$ ,  $(\underline{\lambda}_n)$  is increasing in  $n$ , and  $\lim_{n \rightarrow \infty} \bar{\lambda}_n = \lambda^* = \lim_{n \rightarrow \infty} \underline{\lambda}_n$ . For each  $n \in \mathbb{N}$  and  $\epsilon > 0$ , pick any  $\bar{\pi}_n^\epsilon \in \Pi_\epsilon^*(\bar{\lambda}_n)$  and  $\underline{\pi}_n^\epsilon \in \Pi_\epsilon^*(\underline{\lambda}_n)$ . By definition of  $s(\lambda)$  and the fact that  $\bar{\lambda}_n > \lambda^*$ , we know

$$t(\bar{\pi}_n^\epsilon) \leq T. \quad (4.14)$$

Similarly, from equation (4.13) and the fact that  $\underline{\lambda}_n < \lambda^*$ , we know

$$t(\underline{\pi}_n^\epsilon) \geq T - g_n(\epsilon), \quad (4.15)$$

where  $g_n(\epsilon) = \frac{2\epsilon}{\bar{\lambda}_n - \lambda^*}$ .

We now construct a sequence of randomized policies  $\{\pi_n^\epsilon\}_{n \in \mathbb{N}}$  based on  $\{\underline{\pi}_n^\epsilon\}_{n \in \mathbb{N}}$  and  $\{\bar{\pi}_n^\epsilon\}_{n \in \mathbb{N}}$ . First, we define probabilities  $p_n^\epsilon$  by

$$p_n^\epsilon := \begin{cases} \frac{T - t(\bar{\pi}_n^\epsilon)}{t(\underline{\pi}_n^\epsilon) - t(\bar{\pi}_n^\epsilon)}, & \text{if } t(\underline{\pi}_n^\epsilon) \neq t(\bar{\pi}_n^\epsilon) \text{ and } t(\underline{\pi}_n^\epsilon) \geq T, \\ \frac{1}{2}, & \text{if } t(\underline{\pi}_n^\epsilon) = t(\bar{\pi}_n^\epsilon) \text{ and } t(\underline{\pi}_n^\epsilon) \geq T, \\ 1, & \text{otherwise.} \end{cases} \quad (4.16)$$

The policies are defined by

$$\pi_n^\epsilon(H, U) := \begin{cases} \bar{\pi}_n^\epsilon(H), & \text{if } U > p_n^\epsilon \\ \underline{\pi}_n^\epsilon(H), & \text{if } U \leq p_n^\epsilon. \end{cases} \quad (4.17)$$

Note that all of these policies satisfy  $T \geq t(\pi_n^\epsilon) \geq T - g_n(\epsilon)$ .

We now turn to a technical equality involving the policies  $\{\underline{\pi}_n^\epsilon\}$  and  $\{\bar{\pi}_n^\epsilon\}$ . Let

$$L_n^\epsilon = p_n^\epsilon \underline{\lambda}_n [t(\underline{\pi}_n^\epsilon) - T] + (1 - p_n^\epsilon) \bar{\lambda}_n [t(\bar{\pi}_n^\epsilon) - T]. \quad (4.18)$$

We claim the following equality holds:

$$\lim_{n \rightarrow \infty} \liminf_{\epsilon \rightarrow 0^+} L_n^\epsilon = 0. \quad (4.19)$$

To see this, first note that  $L_n^\epsilon$  can be rewritten as

$$L_n^\epsilon = p_n^\epsilon \bar{\lambda}_n [t(\underline{\pi}_n^\epsilon) - T] + (1 - p_n^\epsilon) \bar{\lambda}_n [t(\bar{\pi}_n^\epsilon) - T] - p_n^\epsilon (\bar{\lambda}_n - \underline{\lambda}_n) [t(\underline{\pi}_n^\epsilon) - T] \quad (4.20)$$

$$= \bar{\lambda}_n [t(\pi_n^\epsilon) - T] - p_n^\epsilon (\bar{\lambda}_n - \underline{\lambda}_n) [t(\underline{\pi}_n^\epsilon) - T]. \quad (4.21)$$

The equality (4.20) can be derived from simple algebra, and equality (4.21) holds due to the definition of  $\pi_n^\epsilon$ . Since  $T \geq t(\pi_n^\epsilon) \geq T - g_n(\epsilon)$  and  $\lim_{\epsilon \rightarrow 0^+} g_n(\epsilon) = 0$  for each  $n \in \mathbb{N}$ , it follows that

$$\liminf_{\epsilon \rightarrow 0^+} \bar{\lambda}_n [t(\pi_n^\epsilon) - T] = 0. \quad (4.22)$$

Similarly, note that

$$\liminf_{\epsilon \rightarrow 0^+} p_n^\epsilon (\bar{\lambda}_n - \underline{\lambda}_n) [t(\underline{\pi}_n^\epsilon) - T] = (\bar{\lambda}_n - \underline{\lambda}_n) \liminf_{\epsilon \rightarrow 0^+} p_n^\epsilon [t(\underline{\pi}_n^\epsilon) - T]. \quad (4.23)$$

Since the  $(p_n^\epsilon)$  and  $[t(\underline{\pi}_n^\epsilon) - T]$  are bounded in  $\epsilon$  the  $\liminf$  above is finite. Since  $\lim_{n \rightarrow \infty} \bar{\lambda}_n - \underline{\lambda}_n = 0$ ,

$$\lim_{n \rightarrow \infty} \liminf_{\epsilon \rightarrow 0^+} L_n^\epsilon = 0 - \lim_{n \rightarrow \infty} (\bar{\lambda}_n - \underline{\lambda}_n) \liminf_{\epsilon \rightarrow 0^+} p_n^\epsilon [t(\underline{\pi}_n^\epsilon) - T] = 0 \quad (4.24)$$

and (4.19) is established.

Now let  $S = \inf_\lambda V(H, \lambda) + \lambda T$ . For any  $n \in \mathbb{N}$  and  $\epsilon > 0$ , because  $S \leq V(H, \underline{\lambda}_n) + \underline{\lambda}_n T$  and similarly for  $\bar{\lambda}_n$ , we have

$$S \leq p_n^\epsilon [V(H, \underline{\lambda}_n) + \underline{\lambda}_n T] + (1 - p_n^\epsilon) [V(H, \bar{\lambda}_n) + \bar{\lambda}_n T] \quad (4.25)$$

$$\leq p_n^\epsilon [r(\underline{\pi}_n^\epsilon) - \underline{\lambda}_n (t(\underline{\pi}_n^\epsilon) - T) + \epsilon] + (1 - p_n^\epsilon) [r(\bar{\pi}_n^\epsilon) - \bar{\lambda}_n (t(\bar{\pi}_n^\epsilon) - T) + \epsilon] \quad (4.26)$$

$$= r(\pi_n^\epsilon) - p_n^\epsilon \underline{\lambda}_n [t(\underline{\pi}_n^\epsilon) - T] - (1 - p_n^\epsilon) \bar{\lambda}_n [t(\bar{\pi}_n^\epsilon) - T] + \epsilon. \quad (4.27)$$

Taking  $\liminf_{\epsilon \rightarrow 0^+}$  on both sides, it follows that

$$S \leq \liminf_{\epsilon \rightarrow 0^+} (r(\pi_n^\epsilon) - p_n^\epsilon \underline{\lambda}_n [t(\underline{\pi}_n^\epsilon) - T] - (1 - p_n^\epsilon) \bar{\lambda}_n [t(\bar{\pi}_n^\epsilon) - T] + \epsilon) \quad (4.28)$$

$$\leq \liminf_{\epsilon \rightarrow 0^+} r(\pi_n^\epsilon) - \liminf_{\epsilon \rightarrow 0^+} (p_n^\epsilon \underline{\lambda}_n [t(\underline{\pi}_n^\epsilon) - T] + (1 - p_n^\epsilon) \bar{\lambda}_n [t(\bar{\pi}_n^\epsilon) - T]). \quad (4.29)$$

Equation (4.19) establishes that taking  $n \rightarrow \infty$  sends the second above term to 0. Hence

$$S \leq \lim_{n \rightarrow \infty} \liminf_{\epsilon \rightarrow 0^+} r(\pi_n^\epsilon). \quad (4.30)$$

We now note that for any policy  $\pi \in \Pi_R$  such that  $t(\pi) \leq T$ , there exists a policy  $\pi' \in \Pi_1(T)$  such that  $r(\pi) = r(\pi')$ . Define the policy  $\pi'$  by letting  $\pi'(H) = \pi(H)$  provided  $\pi(H) \neq \Delta$ . Let  $h \in H_0$  be any point that has already been sampled. Once  $\pi(H) = \Delta$ ,  $\pi'$  chooses to sample at  $h$  for  $\lfloor T - t(\pi) \rfloor$  iterations. Finally,  $\pi'$  chooses to sample at  $h$  one more time with probability  $\lfloor T - t(\pi) \rfloor - (T - t(\pi))$ . By construction it follows that  $t(\pi') = t(\pi)$ , and since sampling at  $h$  does not affect the reward we have that  $r(\pi') = r(\pi)$ .

Since  $T \geq t(\pi_n^\epsilon)$  for all  $n$  and  $\epsilon$ , it follows that  $V_1(T) \geq \sup_{n, \epsilon} t(\pi_n^\epsilon)$ . Thus,

$$\begin{aligned} \sup_{n, \epsilon} r(\pi_n^\epsilon) &\leq V_1(T) \\ &\leq S \\ &\leq \lim_{n \rightarrow \infty} \liminf_{\epsilon \rightarrow 0^+} r(\pi_n^\epsilon) \\ &\leq \sup_{n, \epsilon} r(\pi_n^\epsilon). \end{aligned}$$

Since the first and final terms above are the same, the inequalities above are forced to be equalities, and so  $S = V_1(T)$ .  $\square$

## 5 Experimental Analysis

In this section we run simulations to better understand the behaviour of the optimal policy. We focus on superlevel set detection, and consider two choices for the Markov process  $Y$ . The first is a standard Brownian motion. The second is a compound Poisson process, that is

$$Y(t) = \sum_{i=0}^{N(t)} D_i,$$

where  $N(t)$  is a Poisson process with parameter  $\mu$  and  $D_i$  are independent standard normal variables. We consider the interval  $[a, b] = [0, 1]$ , the threshold  $k = 0$  and assume we have observed endpoint observations  $Y(0) = Y(1) = 0$ . For the compound Poisson process we use a parameter of  $\mu = 20$ . The algorithm we use to compute the value function (and thus the optimal policy) is given in Algorithm 1. Both the standard Brownian motion and compound Poisson process satisfy translation invariance (as defined in equation (3.2)), so we are only concerned with a 3-dimensional state space. To compute the optimal policy we discretize the domain and range of  $Y$ . For all of our experiments we use the indicator reward functions:  $f_+(y) = \mathbb{1}\{y \geq k\}$  and  $f_-(y) = \mathbb{1}\{x \leq k\}$ .

Figure 5.1 depicts the behavior of the optimal policy defined in Algorithm 1 for a Brownian motion over the interval  $[0, 1]$ . The optimal policy and expected value function exhibits several intuitive properties. First, note that the difference between the expected value of sampling and the expected reward of not sampling is larger for the intervals where the endpoints are further away from the threshold. Second, note that the optimal policy takes its first sample exactly in the middle of the interval  $[0, 1]$  at  $x = 0.5$ , the point with the highest variance.

One benefit of being able to compute an optimal policy is being able to characterize suboptimality of the common one-step lookahead heuristic policy described in the introduction. In this problem setup, the one-step lookahead heuristic policy samples the point maximizing the expected immediate reward, or chooses to stop sampling when the gain in reward is lower than the cost. We simulated both this policy and the optimal policy, varying the cost  $c$ . We used both the compound Poisson process and Brownian motion prior on  $Y$ , again over  $[0, 1]$ . As above, we assume the initial history  $H_0 = \{(0, 0), (0, 1)\}$ , that is, we assume initial observations  $Y(0) = Y(1) = 0$ , and use a threshold of  $k = 0$ . The value of a policy is estimated by running the policy under the above conditions (where we sample the observations of  $Y$  from the corresponding conditional distribution), and looking at the expected reward once the policy chooses to stop sampling. We obtain accurate estimates by running each policy 100000 times. The results are shown in Figure 5.2. While the one-step lookahead policy is clearly suboptimal, for each value of  $c$  we tested, the value of the one-step lookahead policy is within 98% of the optimal. This bodes well for the performance of more realistic one-step lookahead algorithms, which are common in practice as discussed in Section 1.

Recall in Section 4 we defined  $V_1(H)$  as the optimal value of a policy with expected budget constraints, and  $V_2(H)$  as the optimal value of a policy with almost sure budget constraints. Also recall we proved  $V_1(H)$  is equal to the solution of the convex optimization program (4.7). In Figure 5.3a we plot  $V_1(H)$  as a function of the expected number of samples  $T$ , when  $Y$  is a Brownian motion, with the same parameters as above. As discussed in Section 4,  $V_1(H) \geq V_2(H)$ . A simple lower bound for  $V_2(H)$  is the one-step lookahead policy that takes exactly  $T$  samples. For each  $T \in \{1, \dots, 10\}$  we estimated this lower bound by simulating the one-step lookahead policy 50000 times. In Figure 5.3b we plot a region containing  $V_2(H)$ : the lower bound is provided by the one-step lookahead policy, and the upper bound is provided by  $V_1(H)$ . The fact that shaded region in Figure 5.3b is small means we have characterized  $V_2(H)$  to high accuracy. The fact that the lower bound provided by the one-step lookahead policy characterizes  $V_2(H)$  to such high accuracy shows that the one-step lookahead policy is very effective in the constrained budget setting.

## 6 Conclusion

In this paper, we consider a class of Bayesian optimization problems where the underlying prior is a Markov process and we pay a cost for each sample. We show that the Bayes-optimal policy is computationally tractable, by way of showing that the value function is completely determined by its values on a 3- or 4-dimensional set. We use this optimal cost-per-sample policy to compute the optimal value when there is

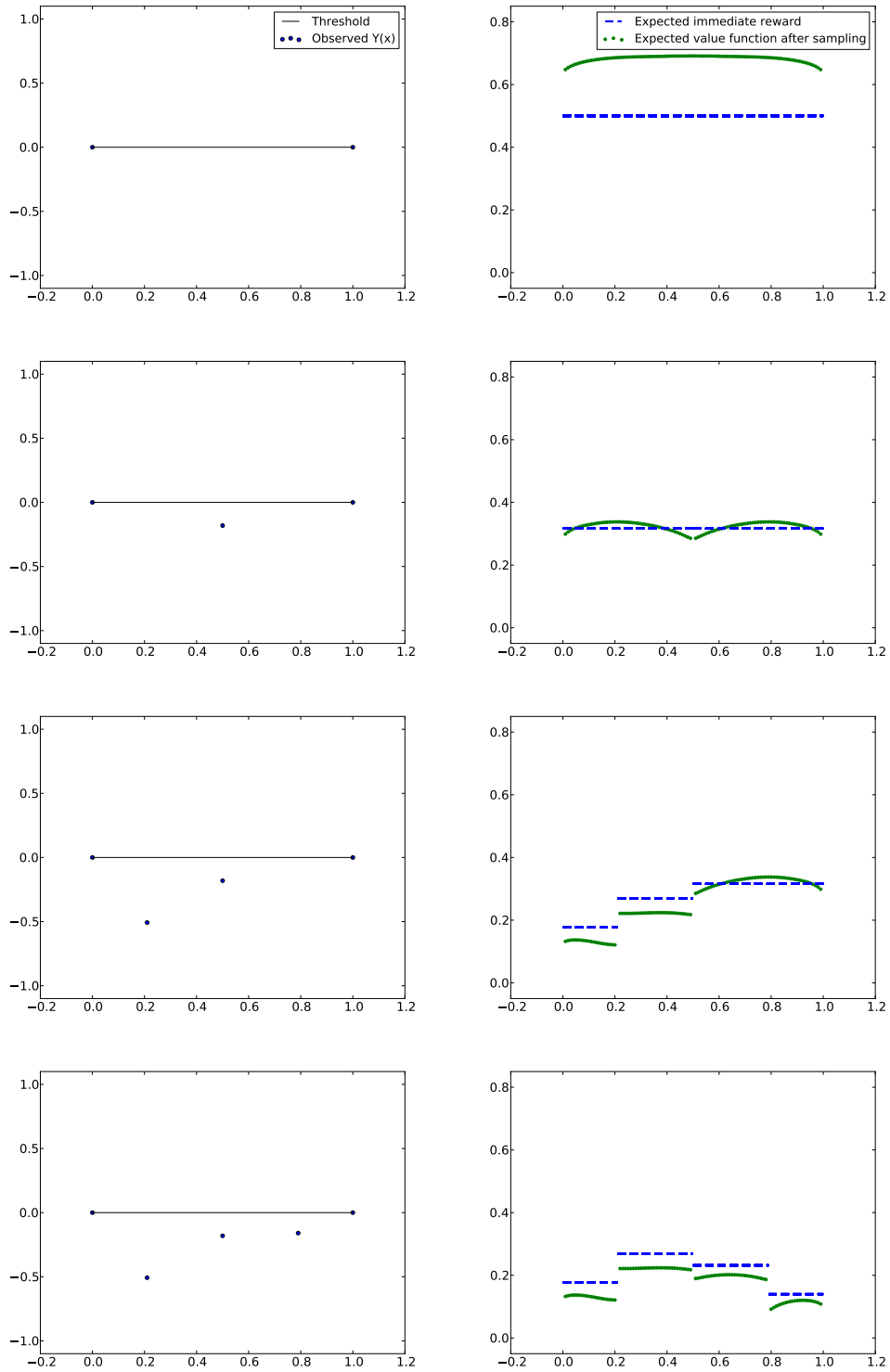
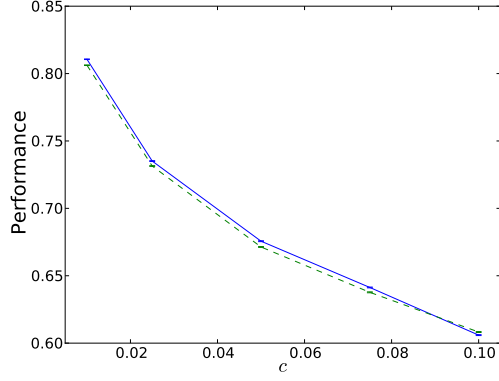
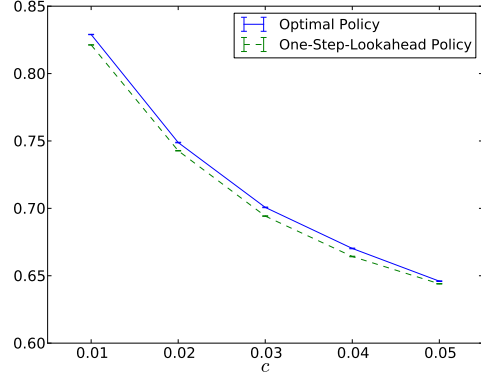


Figure 5.1: Depiction of an optimal policy. Here  $Y$  is a standard Brownian motion, set the threshold  $k = 0$ , x-axis discretization of 100, cost  $c = 0.05$ . **Left:** Sampled points at each iteration. **Right:** Expected value of the value function (solid line) plotted against expected reward (dashed line). The policy samples the point maximizing the difference between these quantities until the maximum difference is negative.

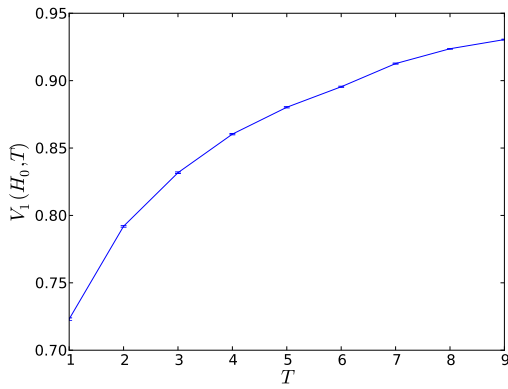


(a)  $Y$  is a Brownian motion.

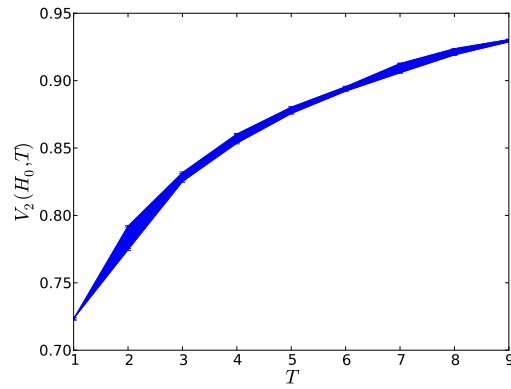


(b)  $Y$  is a compound Poisson process.

Figure 5.2: Value of optimal policy (in blue) and value of one-step lookahead policy (in green) vs. the cost of sampling.



(a) Value of expected constrained budget constrained problems.



(b) Region containing value of exact constrained budget problems.

Figure 5.3: Constrained-budget value plots (expected and almost sure constraints on the left and right respectively) when  $Y$  is a Brownian motion. On the right, the upper bound on  $V_2(H)$  is provided by  $V_1(H)$ , and the lower bound by the one-step lookahead policy that takes exactly  $T$  samples.

no cost to sample, but there is a constraint on the expected number of samples taken, as the result of a simple convex optimization problem. We also use the optimal cost-per-sample policy to provide tight bounds when the constraint on the number of samples taken is almost sure. Computational experiments show that the optimal policy outperforms the commonly used one-step lookahead policy, but also that the optimality gap between one-step lookahead and the optimal policy is small, justifying the use of one-step lookahead in practice.

## A Proofs

**Proof of Lemma 3.** We first show  $V_{[a,b]}(H) \geq V_{[a,b]}(H^I)$ . Let  $\sigma \in \Pi$ . Let  $H^O = H \setminus H^I$  denote the set of initial observations outside  $[a, b]$ . Define  $\pi \in \Pi$  by  $\pi(K) := \sigma(K \setminus H^O)$  for all  $K \in \mathcal{H}$ .

Consider the following Markov processes:

1.  $(H_t^\pi)_{t \geq 0}$  with initial state  $H_0 = H$  operated under  $\mathbb{P}^\pi$ .
2.  $(H_t^{I,\sigma})_{t \geq 0}$  with initial state  $H_0^I = H^I$  operated under  $\mathbb{P}^\sigma$ .

Now, note that

$$R_{[a,b]}(H_t^\pi) = R_{[a,b]}(H_t^\pi \setminus H^O) \approx R_{[a,b]}(H_t^{I,\sigma}) \quad (\text{A.1})$$

where the first equality holds (almost surely, under  $\mathbb{P}^\pi$ ) because  $Y$  is a Markov process and  $H$  contains endpoint observations, and the second equality (in distribution) holds because  $H_t^\pi \setminus H^O$  under  $\mathbb{P}^\pi$  is equal in distribution to  $H_t^{I,\sigma}$  under  $\mathbb{P}^\sigma$ . Moreover,  $\tau^\pi$  under  $\mathbb{P}^\pi$  starting from initial state  $H$  is equal in distribution to  $\tau^\sigma$  under  $\mathbb{P}^\sigma$  starting from initial state  $H^I$ . Thus  $\mathbb{E}^\pi [R_{[a,b]}(H_t^\pi) - c\tau \mid H] = \mathbb{E}^\sigma [R_{[a,b]}(H_t^{I,\sigma}) - c\tau \mid H^I]$ , and the inequality  $V_{[a,b]}(H) \geq V_{[a,b]}(H^I)$  is established.

To establish the reverse inequality, we apply the same logic as above. Let  $\pi \in \Pi$  and define  $\sigma \in \Pi$  by  $\sigma(K) = \pi(K \cup H^O)$ . Then the Markov processes  $(H_t^\pi)_{t \geq 0}$  and  $(H_t^{I,\sigma})_{t \geq 0}$  with initial states  $H$  and  $H^I$  respectively, and operated under  $\mathbb{P}^\pi$  and  $\mathbb{P}^\sigma$  respectively, share the same properties as before, but since  $\sigma$  is now constructed based on  $\pi$  we conclude  $V_{[a,b]}(H^I) \geq V_{[a,b]}(H)$ .  $\square$

**Proof of Lemma 4.** Fix  $H_0 \in \mathcal{H}$ . Since  $\bar{\Pi}_{[a,b]} \subseteq \Pi$  it follows that  $\bar{V}_{[a,b]}(H_0) \leq V_{[a,b]}(H_0)$ . We establish the three properties sequentially:

1. Suppose  $\pi \in \Pi$  is such that  $\mathbb{P}^\pi(\tau = \infty \mid H_0) \neq 0$ . Then  $\mathbb{E}^\pi[\tau \mid H_0] = \infty$ . Since  $c > 0$  it follows that  $\text{Per}(\pi, c, H_0) = -\infty$  and so  $\pi$  cannot be optimal. Thus the supremum can be taken over  $\Pi^1$ .
2. Let  $\pi \in \Pi$  be such that there is nonzero probability that, for some  $t \geq 0$ ,  $\pi(H_t) \in H_t$ . Write  $x = \pi(H_t)$ . Let  $x_1, \dots, x_n$  be the complete set of points that  $\pi$  chooses to sample after sampling  $x$  (note that all of these points are random due to the randomness in the sample, except for  $x_1$ ). Define  $\sigma$  to be the same as  $\pi$ , except on  $H_t$  where  $\sigma$  samples  $x_1$  first and the distribution on the rest of the points is the same. Then  $E^\pi[R_{[a,b]}(H_t)] = E^\sigma[R_{[a,b]}(H_t)]$  since sampling at  $x$  again has no affect on the reward, but  $\sigma$  takes one fewer sample than  $\pi$  and so it has better performance. Repeating this process for every such  $H_t$ , one can construct a policy that never samples a point that has already been sampled and that has better performance than  $\pi$ . It follows that the supremum can be taken over  $\Pi^1 \cap \Pi^3$ .
3. Now, let  $\pi \in \Pi^1 \cap \Pi^3$  and let  $\mathcal{J} = \{H \in \mathcal{H} : \pi(H) \notin [a, b]\}$ . Suppose  $\mathcal{J} \neq \emptyset$  (equivalently,  $\pi \notin \Pi_{[a,b]}^2$ ). For each  $H \in \mathcal{H}$ , let  $x_1, \dots, x_{n_H}$  denote the random sequence of points  $\pi$  samples until it chooses to stop sampling, or it samples inside  $[a, b]$ . That is, if  $\pi(H) \in [a, b]$  then  $n_H=1$ . If  $\pi$  never samples inside  $[a, b]$  after  $H$ , then  $x_{n_H} = \Delta$ . Define  $\sigma(H) = x_{n_H}$ . It follows that  $H_t^\sigma$  with initial state  $H_0$  is equal in distribution to  $H_t^\pi \cap [a, b]$  also with initial state  $H_0$ . Thus  $\mathbb{E}^\sigma[R_{[a,b]}(H_t) \mid H_0] = \mathbb{E}^\pi[R_{[a,b]}(H_t) \mid H_0]$ . However,  $\mathbb{E}^\sigma[\tau \mid H_0] \leq \mathbb{E}^\pi[\tau \mid H_0]$ . Thus, the performance of  $\sigma$  is equal or greater to the performance of  $\pi$ . Hence the supremum can be taken over  $\Pi^1 \cap \Pi_{[a,b]}^2 \cap \Pi^3$  and the result is established.  $\square$

**Proof of Proposition 2.** By Lemma 3 we assume all observations in  $H$  are contained in  $[a, b]$ , i.e.  $x \in [a, b]$  for all  $(x, y) \in H$ .

We show that for any policy  $\pi \in \bar{\Pi}_{[a,b]}$  there exists a policy  $\sigma \in \bar{\Pi}_{[a,b]}$  on  $[a', b']$  such that  $\mathbb{E}^\pi [R_{[a,b]} - c\tau|H] = \mathbb{E}^\sigma [R_{[a',b']} - c\tau|H']$ .

Fix any policy  $\pi \in \bar{\Pi}_{[a,b]}$ . Define  $\sigma \in \bar{\Pi}_{[a',b']}$  by  $\sigma(K) := T_\ell \circ \pi \circ T_{-\ell}(K \cap [a', b'])$  for every  $K \in \mathcal{H}$ . We use the intersection  $K \cap [a', b']$  so that all observations live at or above 0, i.e.  $T_{-\ell}(K \cap [a', b']) \in \mathcal{H}$ .

Now, consider the two Markov processes:

1.  $(H_t)_{t \geq 0}$  under  $\pi$  with initial state  $H_0 = H$ .
2.  $(T_{-\ell}(H'_t))_{t \geq 0}$  under  $\sigma$  with initial state  $T_{-\ell}(H'_0) = T_{-\ell}(H')$ .

In (2), we apply the shift operator  $T_{-\ell}$  to  $H'_t$  so that the two Markov Processes have the same initial state.

We now show the two Markov processes have the same transition kernel. Suppose  $T_{-\ell}(H'_t) = H_t$  for some  $t \geq 0$ , that is, both Markov Processes are in the same state at time  $t$ . Note that  $\pi$  and  $T_{-\ell} \circ \sigma$  choose to sample the same point:

$$T_{-\ell} \circ \sigma(H'_t) = T_{-\ell} \circ T_\ell \circ \pi \circ T_{-\ell}(T_\ell(H_t) \cap [a', b']) = \pi(H_t \cap [a, b]) = \pi(H_t)$$

where the final equality holds because  $H$  has all observations contained in  $[a, b]$  and  $\pi \in \bar{\Pi}_{[a,b]}$ , so  $H_t$  must be contained in  $[a, b]$  for all  $t$ . Call this point  $x_{t+1}$ . It follows that every state  $K$  with nonzero probability for the  $t+1$ th time of both Markov Processes is of the form  $K = H_t \cup \{(x_{t+1}, y)\}$  for some  $y \in \mathbb{R}$ . Then,

$$\mathbb{P}^\pi(H_{t+1} = K | H_t) = \mathbb{P}(Y(x_{t+1}) \in dy | H_t) \quad (\text{A.2})$$

$$= \mathbb{P}(Y(x_{t+1}) \in dy | T_{-\ell}(H'_t)) \quad (\text{A.3})$$

$$= \mathbb{P}^\sigma(T_{-\ell}(H'_t) = K | T_{-\ell}(H'_t)). \quad (\text{A.4})$$

Hence the two Markov Processes have the same transition kernel, and since they have the same initial state, it follows they have the same distribution.

A simple consequence of this is that  $\tau$  under  $\pi$  and  $\tau$  under  $\sigma$  are identically distributed. Indeed,  $\tau_\pi \sim |H_\tau| - |H_0| \sim |H'_\tau| - |H'_0| \sim \tau_\sigma$ .

Finally, observe that the reward is translation invariant:  $R_{[a,b]}(K) = R_{[a',b']}(T_\ell(K))$  for any  $K \in \mathcal{H}$ . Thus,

$$\mathbb{E}^\pi [R_{[a,b]}(H_\tau) | H_0] = \mathbb{E}^\pi [R_{[a',b']}(T_\ell(H_\tau)) | H_0] \quad (\text{A.5})$$

$$= \mathbb{E}^\pi [R_{[a',b']}(T_\ell(H_\tau)) | T_\ell(H_0)] \quad (\text{A.6})$$

$$= \mathbb{E}^\sigma [R_{[a',b']}(H'_\tau) | H'_0] \quad (\text{A.7})$$

where the equality between (A.5) and (A.6) holds because  $T_\ell$  is a bijection, and equality between (A.6) and (A.7) holds because  $T_\ell(H_\tau) | T_\ell(H_0)$  under  $\pi$  is equal in distribution to  $H'_\tau | H'_0$  under  $\sigma$ .

Thus  $\mathbb{E}^\pi [R_{[a,b]} - c\tau | H] = \mathbb{E}^\sigma [R_{[a',b']} - c\tau | H']$ , as we set out to show. It follows that  $V_{[a,b]}(H) \leq V_{[a',b']}(H')$ . Setting  $a := a + \ell$ ,  $b := b + \ell$  and  $\ell := -\ell$  establishes the reverse inequality, and equality follows.  $\square$

## References

- [1] F Archetti and B Betro. A probabilistic algorithm for global optimization. *Calcolo*, 16(3):335–343, 1979.
- [2] B. Betrò and F. Schoen. A stochastic technique for global optimization. *Computers and Mathematics with Applications*, 21(6–7):127–133, 1991.
- [3] Bruno Betrò. Bayesian methods in global optimization. *Journal of Global Optimization*, 1(1):1–14, 1991.

- [4] E Brochu, M Cora, and N de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. Technical Report TR-2009-023, Department of Computer Science, University of British Columbia, November 2009.
- [5] Adam D Bull. Convergence rates of efficient global optimization algorithms. *The Journal of Machine Learning Research*, 12:2879–2904, 2011.
- [6] J Calvin and A Žilinskas. On the convergence of the P-algorithm for one-dimensional global optimization of smooth functions. *Journal of Optimization Theory and Applications*, 102(3):479–495, 1999.
- [7] J Calvin and A Žilinskas. One-Dimensional P-Algorithm with Convergence Rate  $O(n^{-3+\delta})$  for Smooth Functions. *Journal of Optimization Theory and Applications*, 106(2):297–307, 2000.
- [8] J M Calvin and A Zilinskas. One-dimensional Global Optimization Based on Statistical Models. *Non-convex Optimization and its Applications*, 59:49–64, 2002.
- [9] James M. Calvin. A One-Dimensional Optimization Algorithm and Its Convergence Rate under the Wiener Measure. *Journal of Complexity*, 17(2):306–344, June 2001.
- [10] S. E. Chick and P. I. Frazier. Sequential sampling for selection with economics of selection procedures. *Management Science*, 58(3):550–569, 2012.
- [11] E.B. Dynkin and A. A. Yushkevich. *Controlled Markov Processes*. Springer, 1975.
- [12] A Forrester, A Sobester, and A Keane. *Engineering design via surrogate modelling: a practical guide*. Wiley, West Sussex, UK, 2008.
- [13] P. I. Frazier. Tutorial: Optimization via simulation with bayesian statistics and dynamic programming. In C. Laroque, J. Himmelspace, R. Pasupathy, O. Rose, and A. M. Uhrmacher, editors, *Proceedings of the 2012 Winter Simulation Conference Proceedings*, pages 79–94, Piscataway, New Jersey, 2012. Institute of Electrical and Electronics Engineers, Inc.
- [14] P. I. Frazier, W. B. Powell, and S. Dayanik. The knowledge gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613, 2009.
- [15] P.I. Frazier and J. Wang. Bayesian optimization for materials design. arXiv preprint, <http://arxiv.org/pdf/1506.01349.pdf>, 2015.
- [16] Jacob Gardner, Matt Kusner, Zhixiang Xu, Kilian Weinberger, and John Cunningham. Bayesian optimization with inequality constraints. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 937–945, 2014.
- [17] David Ginsbourger and Rodolphe Le Riche. Towards gaussian process-based optimization with finite time horizon. In *mODa 9–Advances in Model-Oriented Design and Analysis*, pages 89–96. Springer, 2010.
- [18] Alkis Gotovos, Nathalie Casati, Gregory Hitz, and Andreas Krause. Active learning for level set estimation. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2013.
- [19] Steffen Grünewälder, Jean-Yves Audibert, Manfred Oppel, and John Shawe-Taylor. Regret bounds for gaussian process bandit problems. In *International Conference on Artificial Intelligence and Statistics*, pages 273–280, 2010.
- [20] Gregory Hitz, Alkis Gotovos, Francois Pomerleau, Marie-Eve Garneau, Cedric Pradalier, Andreas Krause, and Roland Siegwart. Fully autonomous focused exploration for robotic environmental monitoring. In *In Proc. International Conference on Robotics and Automation (ICRA)*, 2014.
- [21] D.R. Jones, M. Schonlau, and W.J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, 1998.



- [22] H. J. Kushner. A new method of locating the maximum of an arbitrary multi-peak curve in the presence of noise. *Journal of Basic Engineering*, 86:97–106, 1964.
- [23] Marco Locatelli. Bayesian algorithms for one-dimensional global optimization. *Journal of Global Optimization*, 10(1):57–76, 1997.
- [24] Marco Locatelli and Fabio Schoen. An adaptive stochastic global optimization algorithm for one-dimensional functions. *Annals of Operations research*, 58(4):261–278, 1995.
- [25] J. Mockus. *Bayesian approach to global optimization: theory and applications*. Kluwer Academic, Dordrecht, 1989.
- [26] C Perttunen and B.E. Stuckman. The rank transformation applied to a multi-univariate method of global optimization. In *Proceedings of the IEEE International Conference on Systems Engineering*, pages 217–220, 1989.
- [27] Cary D Perttunen. A study of alternate stochastic models in kushner-based global optimization methods. In *Systems, Man, and Cybernetics, 1991. Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference on*, pages 597–601. IEEE, 1991.
- [28] Cary D Perttunen and Bruce E Stuckman. The rank transformation applied to a multivariate method of global optimization. *IEEE Transactions on Systems, Man and Cybernetics*, 20(5):1216–1220, 1990.
- [29] W. B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*. John Wiley and Sons, New York, 2007.
- [30] Klaus Ritter. Approximation and optimization on the wiener space. *Journal of Complexity*, 6(4):337–364, 1990.
- [31] M.J. Sasena. *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*. PhD thesis, University of Michigan, 2002.
- [32] Warren Scott, Peter I. Frazier, and Warren B. Powell. The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression. *SIAM Journal on Optimization*, 21(3):996–1026, 2011.
- [33] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012.
- [34] Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 2010.
- [35] Emmanuel Vazquez and Julien Bect. Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *Journal of Statistical Planning and inference*, 140(11):3088–3095, 2010.
- [36] R. Waeber, P. I. Frazier, and S. G. Henderson. Bisection search with noisy responses. *SIAM Journal on Control and Optimization*, 51(3):2261–2279, 2013.
- [37] J. Xie and P. I. Frazier. Sequential bayes-optimal policies for multiple comparisons with a known standard. *Operations Research*, 61(5):1174–1189, 2013.
- [38] Antanas Zilinskas. Axiomatic characterization of a global optimization algorithm and investigation of its search strategy. *Operations Research Letters*, 4(1):35–39, 1985.